

修 士 論 文 の 和 文 要 旨

研究科・専攻	大学院 情報システム学研究科 情報システム設計学専攻 博士前期課程		
氏 名	金城 賀真	学籍番号	0650011
論 文 題 目	領域分割に基づく類似画像検索の高速化		
要 旨			
<p>山本らは、画像の内容に基づく類似画像検索の手法の1つとして、画像を構成する領域の重なり度合いに基づいた、構図による類似画像検索を提案している。画像を領域分割し、各画素にそれが属する領域のラベルを付加した画像(以下、領域画像と呼ぶ)に対し、領域の重なり度合いを測定する尺度として、集合間の類似度である Jaccard 係数をベースとした新しい類似度(重み付き Jaccard 係数に基づく類似度)を提案し、構図に基づく類似画像検索に適用している。</p> <p>しかし、重み付き Jaccard 係数に基づく類似度の計算には、画素が領域画像 I_i, I_j のどの領域に属しているかを1対1に比較しなければならないため、計算量は画素数に対して線形になる。また現状では、1つの問い合わせに対して、データベース内全ての領域画像と類似度を計算しているため、データベース内の領域画像数に比例して検索時間が増大していく。そのため、重み付き Jaccard 係数に基づいた類似画像検索システムを実用的に運用するためには、検索の高速化が必要となる。</p> <p>検索の高速化には、(i)類似度計算時間の削減、(ii)類似度計算回数の削減の2つの手法がある。</p> <p>(i)については、領域画像に対し、runlength 符号化を適用し、同じラベルが連続する画素の部分系列をまとめて表現し、領域のラベルの比較回数を減らし、屢次度計算時間を削減する手法を提案する。領域画像 30,000 枚に対し、runlength 符号化をする場合としない場合の類似度計算時間、領域画像 1 枚当たりのデータサイズを比較した。類似度計算時間は約 1/5、データサイズは約 1/14 に削減されることを確認し、その有効性を示した。また、run の終点の発生分布から、ある位置での領域分割点の発生がベルヌーイ試行でモデル化できると仮定し、その仮定から算出される run の個数の期待値を用いることで、runlength 符号化による計算時間の削減効果について議論した。そのモデルによる類似度計算時間の削減度合いは実測値と近い値であることを確認した。</p> <p>(ii)については、重み付き Jaccard 係数に基づく類似度が距離の公理を満たしていないので、既存の索引構造を適用することができない。そのため、その類似度に対応する新たな索引構造が必要である。索引構造の構築のためには、まず、データの分類が必要であると考え、そのための階層的クラスタリング手法を提案する。あるデータを問い合わせにしたときの他のデータの類似度の順位(ランク)に着目し、お互いを問い合わせにしたとき、お互いのランクが高いデータ組は同じクラスに属するという考えに基づきクラスタリングを行う。具体的には、お互いが k 近傍であるデータ間にエッジを張った k-MNN グラフ(k-Mutually Nearest Neighbor Graph)を定義し、そのエッジ情報に基づきクラスタリングを行う。提案アルゴリズムの評価のために、領域画像 1,000 枚に対し、クラスタリング実験を行った。得られたクラスタ集合と人間によるグループ分けを比較し、人間によるグループ分けと近いクラスタ集合を得ることができることを確認した。また、クラスタ内の類似度の範囲について議論し、クラスタ内の要素間の類似度の値はその他のクラスタの要素への類似度の値に比べて大きいことを確認した。また、大きいデータベース(領域画像 35,782 枚)に対してもクラスタリング実験を行った。そのとき得られたクラスタ集合に、1,000 枚のクラスタリングで得られたクラスタを内包するようなクラスタがある、すなわち、1,000 枚のクラスタリングの結果、同じクラスタであった領域画像集合は、ほとんど、35,782 枚のクラスタリングにおいても同じクラスタに属することを確認した。このことより、提案したクラスタリング手法は、大きなデータベースにも適用可能であることを示した。また、提案したクラスタリング手法を応用したシンプルな索引構造を考え、重み付き Jaccard 係数に基づく類似度に対応した索引構造構築の可能性を示した。</p>			

(大学院情報システム学研究科)